DOCUMENT RESUME

ED 453 289 TM 032 828

AUTHOR Snyder, Patricia
TITLE Treatment of Effect Indices in Journal Editorial Policies: An Editor's Perspective.
PUB DATE 2001-04-14
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Information Analyses (070) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Editors; *Effect Size; *Research Reports; *Scholarly Journals
IDENTIFIERS *Editorial Policy

ABSTRACT

This paper reviews the education and experiences that led one journal editor to support the reporting and interpretation of magnitude-of-effect (ME) indices in substantive research. The paper reviews the controversies associated with the use of ME indices as result interpretation aids and describes the influences of these controversies on journal editorial policies. The role of the editor in ensuring good scientific reporting practices is discussed, and the movements in publication manuals and editorial policies toward routine reporting and interpretation of ME indices are highlighted. Strengths and cautions associated with the routine use of ME indices are reviewed. An appendix lists journals that have adopted ME index requirements. (Contains 46 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Running Head: TREATMENT OF EFFECT INDICES IN JOURNALS

ED 453 289

Treatment of Effect Indices in Journal Editorial Policies:

An Editor's Perspective

Patricia Snyder

Louisiana State University Health Sciences Center

TM032828

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. Snyder

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA, April 14, 2001.

BEST COPY AVAILABLE

Abstract

In this paper, I review the education and experiences that led one journal editor to support the reporting and interpretation of magnitude-of-effect (ME) indices in substantive research. I review the controversies associated with the use of ME indices as result interpretation aids and describe the influences of these controversies on journal editorial policies. The role of the editor in ensuring good scientific reporting practices is discussed and the movements in publication manuals and editorial policies toward routine reporting and interpretation of ME indices are highlighted. Strengths and cautions associated with the routine use of ME indices are reviewed.

Treatment of Effect Indices in Journal Editorial Policies: An Editor's Perspective

I appreciate the opportunity to share my perspectives related to the treatment of effect indices in journal editorial policies. To begin, I will share how my thinking about effect indices has been shaped during my career to date. I will then describe what I view as major areas of agreement and disagreement among research methodologists related to the reporting of effect indices and how editorial practices are influenced accordingly. The extent to which authors and editors are following the "encouragement" to report effect indices provided in the fourth edition of the Publication Manual of the American Psychological Association (APA, 1994) will be discussed. I will describe reasons why the editor's role is crucial for ensuring good scientific reporting practices, including reporting effect indices. A review of journals that are requiring the reporting and interpretation of effect indices will be provided. Finally, I will discuss issues likely to arise when journal editors require the reporting and interpreting of effect indices, given strengths and cautions associated with ME indices.

Influences on My Perspective

I believe magnitude-of-effect (ME) indices are important result interpretation aids and should be used by researchers to help them evaluate the importance of a difference or a relationship. Unlike the journal editors described by Hyde (2001) in her thought-provoking article related to the roles of editors, textbook authors, and publication manuals in relation to effect index reporting, I was not socialized into my "statistical morality" 30 or more years ago. My socialization related to the importance of ME indices occurred during my doctoral training in the late 1980s and early 1990s.

My perspective related to the importance of employing additional result interpretation aids in quantitative investigations, beyond tests of statistical significance, was heavily influenced by the writings of Carver (1978), Cohen (1990), Rosnow and Rosenthal (1989), and Kupfersmid (1988) and the teaching and writings of Bruce Thompson (e.g., Thompson, 1988, 1989). I was fortunate to begin my doctoral training during a time when a much publicized round of written and oral debates were occurring in psychology and education related to the strengths and limitations of statistical significance testing. As Rosnow and Rosenthal (1989) noted, much of what was being said during this time had been said before, however, it was important that we hear it all again, so that my generation and I would be aware of the potential pitfalls of statistical inference and recognize viable alternatives or supplements.

In 1992, I presented a paper at the American Educational Research Association conference with my colleague Steve Lawson in which we (a) described why methodologists encourage the use of ME indices, and (b) reviewed different types of ME estimates. This paper (Snyder & Lawson, 1993) subsequently was published in a special issue of the Journal of Experimental Education titled "Statistical Significance Testing in Contemporary Practice." After completing a comprehensive literature review for this paper, I became even more convinced about the usefulness of ME indices as result interpretation aids; despite my understanding of cautions associated with the use of these measures (e.g., O'Grady, 1982). I also began to believe that user-friendly descriptions of the myriad of ME indices and how to select among these options should be made widely available. Otherwise, researchers in my field, who are excellent substantive researchers, but not research methodologists, would be unlikely to move beyond reporting ME indices that are routinely produced in common statistical packages (cf. Kirk, 1996). Because of my early experiences, I became committed to ensuring that, when appropriate,

papers I wrote would report and interpret ME estimates. I also decided that if I became involved as a reviewer or editor of manuscripts I would share my perspective about the potential usefulness of ME indices as result interpretation aids with those in my substantive field.

I have served on the editorial boards of four journals related to my substantive and methodological interests since 1995. As I have conducted what I estimate to exceed 350 manuscript reviews over approximately 6 years, I have seen very few instances of authors reporting and interpreting ME indices on first submission. However, I have read many manuscripts in which authors interpreted p calculated values as "highly" significant or "approaching" statistical significance. I have encountered language in manuscripts that implied many authors were confused about what statistical significance tests do and do not tell us (e.g., do not inform about result importance or result replicability). In some instances, these "misinterpretations" affected the validity of inferences drawn by the authors, in other cases these errors were less consequential due to the strengths of the study design and methods employed. Nevertheless, I routinely request in my reviews that authors report and interpret ME estimates and I briefly provide a rationale for this request. Frequently, I offer references for authors to locate information about ME alternatives. My review experiences have strengthened my belief that researchers typically should report and interpret ME estimates to "shore up facts and inductive inferences" (Rosnow & Rosenthal, 1989, p. 1276).

In 1997, I became an Associate Editor for the Journal of Early Intervention (JEI), the leading scholarly journal in a field concerned with services and supports to young children with special needs, their families, and the personnel who serve them. The Editor of JEI, R.A. McWilliam, a contemporary of mine in relation to his statistical socialization, shares my belief about the usefulness of ME estimates as result interpretation aids. After engaging in discussions

with members of the JEI editorial board, other researchers in our field, and colleagues from other specialty areas in psychology and education, Dr. McWilliam made a decision to publish a series of editorial guidelines for JEI authors and reviewers. These guidelines are not intended to be "editorial policing" (cf. Robinson & Levin, 1997). Rather, they are designed to inform authors and reviewers about how work submitted to JEI will be judged, beyond criteria stated in the fourth edition of the Publication Manual of the American Psychological Association (APA, 1994).

Because work submitted to JEI encompasses a variety of research traditions (e.g., group quantitative, single-case experimental designs, qualitative), we believe authors, reviewers, and readers of the journal should have an explicit understanding of editorial policies related to these traditions. I authored the guidelines for "group quantitative" investigations after obtaining significant input from authors and reviewers in the field (Snyder, 2000). Among other points, these guidelines contain information about (a) the importance of ME estimates as result interpretation aids; (b) the types of ME indices available to researchers, with supporting references provided; and (c) why we are requiring that authors who submit manuscripts to JEI report and interpret ME estimates.

My experiences to date as a graduate student, author, reviewer, and associate editor are not necessarily unique and my review of these experiences is not intended to be self-serving. I have reviewed the events that have helped shaped my thinking about ME indices because they demonstrate that, over time, education and experience can influence editorial policies. As Thompson (1999b) noted, the field moves, albeit slowly. Continued education about good statistical practices should ensure that ME indices receive favorable treatment in journal editorial policies.

To Report or Not to Report ME Indices: Areas of Agreement and Disagreement

The "controversies" about whether authors should report or editors should request ME indices are intertwined with ongoing debates related to the use of null hypothesis statistical significance testing (NHST). And, as Kirk (1996) noted, for almost 70 years, null hypothesis significance testing has been surrounded by controversy.

One group of methodologists has suggested that educational and psychological research would be better off without NHST or nil null hypothesis testing (Cohen, 1994), and have issued calls to abandon these practices (e.g., Berkson, 1938; Carver, 1978; Cohen, 1900; 1994; Meehl, 1978; Schmidt, 1996; Schmidt & Hunter, 1997; Rosenboom, 1960; 1997). These individuals have suggested that NHST should be replaced (in single studies) with other result interpretation aids (e.g., ME indices, confidence intervals, or Bayesian methods of inference).

Another group of methodologists believes that although NHST has been misunderstood and misused, the practice should not be abandoned (e.g., Abelson, 1997; Chow, 1988; Levin 1993; Levin & Robinson, 1999; Mulaik, Raju, & Harshman, 1997; Robinson & Levin, 1997). These methodologists have suggested a variety of strategies for using statistical hypothesis tools more appropriately, using what Levin (1998, p. 329) has labeled "intelligent statistical hypothesis testing alternatives."

Both the proponents and detractors of NHST generally appear to agree that ME indices can serve useful functions in the interpretation of substantive findings, though enthusiasm for mandatory reporting and interpretation of these measures appears to vary inversely in relation to support for NHST practices. Those who support NHST often suggest that ME indices are important supplements to, but not replacements for, tests of statistical significance. For example, Robinson and Levin (1997) have suggested a two-step approach to the reporting and evaluation

of empirical results where evaluation of the magnitude and substantive significance of obtained effects is conditional upon statistical significance. These authors noted that, "Researchers cannot live by effect sizes alone!" (p. 23). Alternatively, those who do not support NHST show much greater support for ME indices. Carver (1993), for example, stated,

> Statistical significance testing tells us nothing directly relevant to whether the results we found are large or small, and it tells us nothing with respect to whether the sampling error is large or small. We can eliminate this problem by reporting both effect size and standard errors. (p. 291)

Nickerson (2000) observed that nobody, to his knowledge, has argued that NHST is the only type of analysis of data that one needs to perform. Similarly, I have not read an article advocating the use of a ME point estimate as a sole result interpretation aid. Most often, those who support the reporting and interpretation of ME indices advocate for the inclusion of confidence intervals (e.g., Kirk, 1996; Wilkerson and the Task Force on Statistical Inference, 1999) or other aids.

Given the controversies described and the areas of agreement and disagreement related to ME indices, what are contemporary journal editors to do in relation to formulating editorial policy? Should they rely solely on guidance offered in publication manuals, or, should they formulate additional policies? Nickerson (2000), for example, was motivated to explore the controversies associated with NHST partly because of his efforts to develop a policy that would ensure the journal he edited did not publish articles reflecting egregious misuses of the method. He commented that his in-depth exploration of the controversies surrounding NHST did not lead him directly to understanding its proper role in social science research. He suggested that a more likely consequence of extensive review of the controversies surrounding NHST, including the

reporting and interpretation of ME indices, is the discovery that some principles and relationships considered well-established or taken for granted are not beyond dispute. He concluded that statistical methods should facilitate good thinking, and only to the degree that they do so are they being used well.

I agree with Nickerson, particularly as regards intelligent use of ME indices. To borrow slightly from Levin (1998), I believe editors should practice "intelligent scientific editing." They should move beyond "encouraging" authors to provide information about ME. The editorial policy promulgated by Murphy (1997), editor of the Journal of Applied Psychology, makes the relevant point clearly:

> If an author decides not to present an effect size estimate along with the outcomes of a significance test, I will ask the author to provide specific justification for why effect sizes are not reported. So far, I have not heard a good argument against presenting effect sizes. Therefore, unless there is a real impediment to doing so, you should routinely include effect size information in the papers you submit. (p. 4)

Have Encouragements to Provide ME Indices Changed Practices?

I remember the enthusiasm I felt when the fourth edition of the Publication Manual of the American Psychological Association (APA, 1994) was published. I was particularly pleased to see inclusion of information related to the two types of probability values; the requirements to report sufficient statistics; the mandate to provide alpha levels and exact p calculated values; and the encouragement to provide effect-size information, since "neither of the two types of probability values reflects the importance (magnitude) of an effect or the strength of a relationship . . . " (p. 18).

Shortly after the publication of the fourth edition, I had a conversation with Bruce Thompson about the revised manual. We contemplated whether the encouragement to provide effect size information would be sufficient to lead to changes in practices by authors. These conversations led to a series of studies we conducted (Snyder & Thompson, 1999; Thompson & Snyder, 1997, 1998) to evaluate empirically the impact of the APA "encouragement."

Our findings and those of others (e.g., Kirk, 1996; Vacah-Haase & Nilsson, 1998; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000) confirmed that encouragement to provide ME information has had minimal impact on ME reporting and interpretation practices. Kirk (1996), for example, examined the 1995 volumes of four APA journals and found that the percentage of articles that included a ME estimate ranged from 12% for the Journal of Experimental Psychology to 77% for the Journal of Applied Psychology. He noted the better showing for the latter journal may be somewhat misleading and he attributed this finding to more frequent use of regression and correlation procedures in this journal. Because computer packages routinely provide ME measures for these procedures (e.g., R^2 and "adjusted" R^2), authors may report these indices, but they do not necessarily interpret them.

Across all four journals Kirk reviewed, R^2 and the coefficient of determination accounted for 60% of the ME indices reported. Omega squared, intraclass correlation coefficients, and Cohen's f, which Kirk stated would be more appropriate for the analysis of variance procedures typically used in the Journal of Experimental Psychology, were rarely reported (less than .05% of the ME indices reported). Kirk insightfully suggested these less common ME indices may be reported infrequently because they are not part of the "default" reports provided in computer printouts for ANOVA procedures.

Even those who are somewhat less enthusiastic about the necessity for ME reporting and interpretation acknowledge the APA encouragement has not resulted in significant changes in reporting practices related to these indices (e.g., Levin & Robinson, 1999). Levin and Robinson, however, have expressed concerns related to editorial policies moving beyond encouragement to requiring authors to report ME indices. They stated such policies might result in authors who feel "sanctioned" in restricting their reporting to ME indices only. I would be surprised to find a journal editor who would accept an article for publication that included a ME point estimate as the sole result interpretation aid. As stated previously, even the harshest critics of NHST favor the reporting of both ME point estimates and confidence intervals in individual studies, and meta-analyses in the integration of multiple studies (e.g., Schmidt, 1996).

Not surprisingly, I find myself agreeing with Thompson (1999b) that the encouragement to provide ME indices generally has been ineffective and editors should either require these indices or ask authors to provide specific justification for why they are not reported and interpreted. Furthermore, I believe editors should heed the advice of Kirk (2001) who stated, "Promoting the reporting of measures of effect magnitude is important, but that is only part of a much larger issue of promoting good statistical practices" (p. 215).

The Role of the Editor in Promoting Good Statistical Practices

Journal editorial policies are crucial for promoting good statistical practices. Glantz (1980) suggested that journals are the major force for quality control in scientific work. Kirk (2001) characterized journal editors as the gatekeepers for what appears in scientific journals. He noted editors have a responsibility to be knowledgeable about good statistical practices and to make authors follow those practices. Kirk (1996) suggested that significant modifications to journal editorial policies could set off chain reactions. Statistics teachers change courses,

textbook authors revise statistics books, and journal authors modify their practices. Unfortunately, controversies continue to abound related to what constitutes good statistical practices and how they should be reflected in journal editorial policies (cf. Nickerson, 2000).

Attempts to define good statistical reporting practices have been made (e.g., APA, 1994; Wilkinson & the Task Force on Statistical Inference, 1999). As regards ME indices, Hyde (2001) recently suggested that reporting these indices is a minimum scientific standard. Wilkinson and the Task Force on Statistical Inference stated, "reporting and interpreting effect sizes in the context of previously reported effects is essential to good research" (p. 599). They noted ME indices should always be presented for primary outcomes and authors should add brief comments to place these indices in practical and theoretical contexts. These authors also affirmed that interval estimates should be given for any ME index involving principal outcomes. Most important, Wilkinson et al. presented their recommendations related to ME reporting and interpretation in a context that acknowledges statistical conclusion validity as only one aspect of design validity that editors should consider.

A number of journal editors also endorse the premise that reporting and interpreting ME indices reflect good statistical practices. They have promulgated editorial policies that "require" rather than "encourage" authors to report and interpret these indices. Appendix A shows a list of 14 journals with policies related to reporting ME estimates. My review of these policies revealed no editor suggested ME indices should be the sole result interpretation aid used in substantive research. In fact, the majority of these policies suggest ME indices should supplement inferential tests of statistical significance (e.g., Ellis, 2000; McLean & Kaufman, 2000).

The ongoing evolution in editorial policies provides a basis for cautious optimism related to routine reporting and interpretation of ME indices (Thompson, 1999b). Based on the

recommendations of Wilkinson and the Task Force on Statistical Inference (1999) and the positive endorsement for ME indices by the chair of the Publications and Communications Board of the American Psychological Association, which supervises the revision of the Publication Manual (Hyde, 2001), it appears likely that the 5th edition of the manual may move beyond "encouraging" reporting of ME indices. This movement should be accompanied, however, by ongoing efforts to educate editors and authors about issues associated with ME indices. Otherwise, as Kirk (2001) noted, we run the risk of having editors and researchers blindly adhering to ME reporting practices much as they may be blindly adhering to NHST.

Issues Likely to Arise when Editorial Policies "Require" ME Reporting and Interpretation

Editors, reviewers, and authors must be informed about the strengths and cautions associated with ME measures. As Nickerson (2000, p. 281) observed, "One practical impediment to the use of effect-size indicants . . . may be poor understanding of them among researchers." Thus, I believe editorial guidelines or policies should continue to play an important role in ensuring that messages about the usefulness and limitations of ME indices in relation to the justification of knowledge are heard, beyond whatever information is contained in the next edition of the Publication Manual.

Beyond providing interpretative aid related to result importance in a single study, Thompson (2000) noted routine reporting of ME indices is useful for at least three reasons. First, meta-analytic work will be facilitated. Second, reporting ME indices creates a literature base that enables researchers to formulate specific study expectations easily, by integrating effects reported in previous studies. Third, interpreting ME indices in a given study facilitates evaluation of (a) how results from one study fit into the existing literature, (b) how similar or dissimilar results are across related studies, and (c) what study features contributed to ME indices.

Despite the strengths of ME indices as result interpretation aids, a number of issues are likely to arise when the reporting and interpreting of ME indices becomes more widespread. Editors are likely to receive inquiries from authors about what types of ME indices should be reported and interpreted. Clarification will continue to be needed about the various types of ME indices, how these indices are conceptually related, and when they might be appropriately used (e.g., Kirk, 1996; Olejnik & Algina, 2000; Rosenthal, 1994; Snyder & Lawson, 1993; Thompson, 1999a). Investigations of how various ME indices perform under real and simulated conditions will be important to help editors provide informed guidance (e.g., Yin & Fan, 2001).

Because interpretation of the noteworthiness of ME indices is subjective, editors may find themselves reaching different conclusions regarding the noteworthiness of effects than authors. Sometimes a small effect may be very important, other times a larger effect may not be noteworthy. To support subjective judgments about the noteworthiness of findings, editors will need to highlight the necessity for authors to interpret their effects in the context of the study design (e.g., relationships between categories of the independent variables, reliability of dependent measure scores for study participants) and in relation to similar studies. Editors may be tempted to adopt Cohen's (1988) guidelines for what constitutes small, medium, and large effects. Setting arbitrary guidelines against which to evaluate the size of a particular ME discounts the context dependency of the investigative process (Snyder & Lawson, 1993). As Thompson (1999a, p. 34) noted, if we apply Cohen's conventions with the same rigidity that we have traditionally applied to alpha = .05 in NHST, we will merely be "stupid" in a new metric.

Promoting understanding of the context-dependency issues associated with ME estimates is critical. In fixed-effect design models, statistical generalization is impossible for levels not included in the design. In these instances, editors should not permit authors to state the

percentage of variance accounted for in the dependent variable by the independent variable is 30%. A more accurate statement would be that k particular levels of the independent variable accounted for 30% of the variance in the dependent variable when n subjects of p type were assigned to each cell. Researchers can choose levels of treatment known to vary widely and increase the probability of obtaining a large value for the index of association strength. The addition of variables in a study that will include a multiple regression analysis, for example, may increase the value of R^2, simply due to sampling error variance. Editors should request that authors report adjusted or corrected ME estimates in these instances. Further, they should help authors understand what design features contribute to sampling error variance.

Finally, editors and publication manuals must help authors understand that ME estimates are point estimates and that confidence intervals can, and probably should, be constructed as part of ME reporting practices (Wilkinson & Task Force on Statistical Inference, 1999). As noted in the recommendations of the Task Force, comparing confidence intervals from a current study to intervals found in previous studies will help focus attention on stability across studies.

One of the responsibilities and privileges of being a journal editor is communicating specific statements about the kind of methodological and statistical quality the editor is striving for in manuscripts published in the journal, particularly in light of developments in publication policies related to good scientific reporting practices (cf. Levin, 1993). Editors who recognize the usefulness and cautions associated with ME statistics and other result interpretation aids are more likely to author informed guidelines and treat ME constructively in their policies. In the final analysis, ME indices and other result interpretation aids are merely tools researchers use to help them gain a more informed analysis of data. To paraphrase Thompson (1997), we should avoid letting these result-interpretation-aid "tails" wag the dog of sound scientific inquiry.

References

Abelson, R. P. (1997). The surprising longevity of flogged horses: Why there is a case for the significance test. Psychological Science, 8, 12-15.

American Psychological Association (1994). Publication manual of the American Psychological Association (4th ed.). Washington, DC: Author.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. Journal of the American Statistical Association, 33, 526-542.

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Carver, R. P. (1993). The case against statistical significance testing, revisited. Journal of Experimental Education, 61, 287-292.

Chow, S. L. (1988). Significance test or effect size? Psychological Bulletin, 103, 105-110.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45, 1304-1312.

Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49, 997-1003.

Ellis, N. (2000). Editorial. Language Learning, 50 (3).

Glantz, S. A. (1980). Biostatistics: How to detect, correct, and prevent errors in the medical literature. Circulation, 61, 1 –7.

Hyde, J. S. (2001). Reporting effect sizes: The roles of editors, textbook authors, and publication manuals. Educational and Psychological Measurement, 61, 225-228.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. Educational and Psychological Measurement, 56, 746-759.

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. Educational and Psychological Measurement, 61, 213-218.

Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. American Psychologist, 43, 635-642.

Levin, J. R. (1993). Statistical significance testing from three perspectives. Journal of Experimental Education, 61, 378-382.

Levin, J. R. (1998). To test of not to test Ho? Educational and Psychological Measurement, 58, 313-333.

Levin, J. R., & Robinson, D. H. (1999). Further reflections on hypothesis testing and editorial policy for primary research journals. Educational Psychology Review, 11, 143-155.

McLean, J. E., & Kaufman, A. S. (2000). Editorial: Statistical significance testing and Research in the Schools. Research in the Schools, 7, (2).

Meehl, P.E. (1978). Theoretical risk and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Mulaik, S. A., Raju, N. S., & Harshman, R. A. (1997). There is a time and place for significance testing. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), What if there were no significance tests? (pp. 65-115). Mahwah, NJ: Erlbaum.

Murphy, K. R. (1997). Editorial. Journal of Applied Psychology, 82, 3-5.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. Psychological Methods, 5, 241-301.

O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. Psychological Bulletin, 92, 766-777.

Olejnik, S., & Algina, J. (2000). Measure of effect size for comparative studies: Applications, interpretations, and limitations. Contemporary Educational Psychology, 25, 241-286.

Robinson, D. H., & Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. Educational Researcher, 26 (5), 21-26.

Rosenboom, W. W. (1960). The fallacy of the null hypothesis significance test. Psychological Bulletin, 57, 416-428.

Rosenboom, W. W. (1997). Good science is abductive, not hypothetico-deductive. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), What if there were no significance tests? (pp. 335-392). Mahwah, NJ: Erlbaum.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L.V. Hedges (Eds.), The handbook of research synthesis (pp. 231-244). New York: Russell Sage Foundation.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. Psychological Methods, 1, 115-129.

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), What if there were no significance tests? (pp. 37-64). Mahwah, NJ: Erlbaum.

Snyder, P. (2000). Guidelines for reporting results of group quantitative investigations. Journal of Early Intervention, 23, 145-150.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61, 334-349.

Snyder, P., & Thompson, B. (1998). Use of tests of statistical significance and other analytic choices in a school psychology journal: Review of practices and suggested alternatives. School Psychology Quarterly, 13, 335-348.

Thompson, B. (1988, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY (ERIC Document Reproduction Service No. ED 301 595).

Thompson, B (1989). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling and Development, 22, 2-5.

Thompson, B. (1999a, April). Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. Invited address presented at the annual meeting of the American Educational Research Associations, Montreal.

Thompson, B. (1999b). Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. Educational Psychology Review, 11, 157-169.

Thompson, B. (2000, June). A suggested revision to the forthcoming 5th edition of the APA Publication Manual [On-line]. Available: http//acs.tamu.edu/~bbt6147/apaeffec.htm

Thompson, B., & Snyder, P. (1997). Statistical significance testing practices in the Journal of Experimental Education. Journal of Experimental Education, 66, 75-83.

Thompson, B., & Snyder, P. (1998). Statistical significance and reliability analyses in recent JCD research articles. Journal of Counseling and Development, 76, 436-441.

Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T.S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. Theory and Psychology, 10, 413-425.

Vacha-Haase, T., & Nilsson, J. E. (1998). Statistical significance reporting: Current trends and usages within MECD. Measurement and Evaluation in Counseling and Development, 31, 46-57.

Wilkinson, L. & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 594-604. [Available: http://www.apa.org/journals/amp/amp548594.html]

Yin, P., & Fan, X. (2001). Estimating R^2 shrinkage in multiple regression: A comparison of different analytic methods. Journal of Experimental Education, 69, 203-224.

Appendix

Journals that Have Editorial Policies "Requiring" Magnitude-of-Effect (ME) Reporting and Interpretation

Career Development Quarterly

Contemporary Educational Psychology

Exceptional Children

Educational and Psychological Measurement

Journal of Agricultural Education

Journal of Applied Psychology

Journal of Consulting and Clinical Psychology

Journal of Early Intervention

Journal of Experimental Education

Journal of Learning Disabilities

Language Learning

Measurement and Evaluation in Counseling and Development

The Professional Educator

Research in the Schools